

Communiqué de presse – 03 mars 2025

Avancée majeure en biologie : une IA générative hybride conçoit de nouvelles molécules

Prédire la structure des protéines, en créer de nouvelles. Le design des protéines est un domaine de recherche en pleine révolution grâce aux possibilités offertes par l'IA, avec en ligne de mire la conception de nouvelles protéines et enzymes avec des applications directes en santé et en environnement. Retour sur ces avancées par des chercheurs en mathématiques d'INRAE qui ont développé une IA combinant apprentissage et raisonnement. Cette IA hybride est ainsi capable de concevoir des protéines selon des règles apprises par apprentissage profond, mais aussi issues de la physique ou explicitées par les designers. Des résultats qui contribuent au panorama des méthodes de design de protéines présentées dans la revue *Nature Reviews Methods Primers*.

Les protéines jouent un rôle central dans le développement et le fonctionnement des êtres vivants. Chimiquement, une protéine déroulée ressemble à un collier de perles, chaque perle étant un acide aminé (il en existe 20 différents dans le vivant). En fonction du type et du nombre d'acides aminés, la protéine acquiert une forme spécifique dans l'espace. Ce phénomène de « repliement » détermine la fonction de la protéine, comme le transport de molécules dans le sang, la digestion enzymatique, la réception et transmission de signaux. Au final, notre organisme rassemble près de 20 000 types de protéines différents. Le passage d'une séquence d'acides aminés à une forme 3D est un phénomène très complexe, et sa compréhension constitue un enjeu de recherche important, notamment pour mieux étudier certaines maladies causées par un mauvais repliement des protéines, dont des maladies neurodégénératives comme Alzheimer et Parkinson.

Actuellement, un domaine en recherche progresse à une vitesse phénoménale : le design de protéines par ordinateur, mis récemment en avant par le prix Nobel de chimie 2024. La capacité de concevoir des protéines totalement nouvelles, pour un objectif bien précis, ouvre en effet des possibilités révolutionnaires dans le domaine de la santé et de l'environnement.

INRAE, au travers d'une collaboration durable entre chercheurs en intelligence artificielle et modélisateurs moléculaires¹, a contribué à ces avancées au travers de méthodes de design basées sur une intelligence artificielle hybride, mêlant apprentissage et raisonnement, déjà mise en œuvre pour concevoir différentes nouvelles protéines, fonctionnelles, et caractérisées expérimentalement.

Historiquement, la physique avait permis de réduire, de façon partiellement satisfaisante, le problème de design de protéines à une problématique mathématique d'optimisation difficile. Mais l'IA a bousculé cette approche. La composition chimique d'une protéine étant définie par un texte simple (sa séquence) qui définit aussi sa forme 3D finale, les IA génératives du langage naturel (ChatGPT, Gemini ou Llama) ont été immédiatement adaptées au langage des séquences de protéines naturelles et celles ciblées sur la génération d'images en 2D (DALL-E, Midjourney ou Flux) étendues pareillement pour générer de nouvelles structures de protéines. La difficulté principale est de guider ces IA pour créer la protéine qui a les capacités souhaitées.

¹ Notamment à travers l'implication dans les projets de l'ANITI (Artificial and Natural Intelligence Toulouse Institute).

Une IA hybride capable d'apprendre à jouer à des jeux logiques comme le sudoku par la simple observation de grilles résolues

Les méthodes développées par les chercheurs d'INRAE combinent pour cela 2 familles d'outils IA : l'apprentissage profond et le raisonnement automatique.

L'apprentissage profond est utilisé pour extraire les règles qui régissent la conception de protéines, en exploitant les séquences et structures de protéines naturelles, accumulées depuis des décennies par les biophysiciens dans la Protein Data Bank. Le raisonnement automatique est utilisé pour sa capacité à combiner ces règles apprises avec des lois fondamentales de la physique ou des directives du concepteur, pour identifier, très rapidement, les différentes protéines qui répondent à ces exigences parmi l'univers exponentiel des protéines possibles.

L'ensemble forme une IA générative, dite neuro-symbolique, capable de concevoir des protéines respectant précisément les consignes des designers. Les chercheurs ont montré que cette architecture est aussi capable d'apprendre à jouer parfaitement à des jeux logiques, tels que le sudoku, sans qu'on en explicite les règles, juste en observant des grilles résolues.

Ces méthodes démocratisent considérablement la capacité de concevoir de nouvelles protéines, même si la maîtrise du design est un chemin qui nécessite encore de nombreuses recherches.

Référence

Albanese K.I., Barbe S., Tagami S. et al. (2025). Computational Protein Design. *Nature Reviews Methods Primers*, DOI : <https://doi.org/10.1038/s43586-025-00383-1>

Contacts scientifiques :

Thomas Schiex - thomas.schiex@inrae.fr

Unité de recherche Mathématiques et informatique appliquées Toulouse

Département scientifique Mathématiques, informatique, sciences de la donnée et technologies du numérique (MATHNUM)

Centre INRAE Occitanie-Toulouse

Sophie Barbe - sophie.barbe@inrae.fr

Unité mixte de recherche Toulouse Biotechnology Institute

Départements scientifiques Microbiologie et chaîne alimentaire (MICA) et Sciences pour l'ingénierie des aliments, des produits biosourcés et des résidus de l'activité humaine (TRANSFORM)

Centre INRAE Occitanie-Toulouse

Contact presse :

Service Médias et opinion INRAE: 01 42 75 91 86 – presse@inrae.fr

INRAE, l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, est un acteur majeur de la recherche et de l'innovation. L'institut rassemble une communauté de plus de 10 000 personnes, dont 8000 personnels permanents et plus de 2500 contractuels financés sur projet chaque année, avec plus de 270 unités de recherche, de service et d'expérimentation implantées dans 18 centres sur toute la France.

Institut de recherche finalisée, il se positionne parmi les tout premiers organismes de recherche au monde en sciences agricoles et alimentaires, en sciences du végétal et de l'animal, et en écologie-environnement. Il est le premier organisme de recherche mondial spécialisé sur l'ensemble « agriculture-alimentation-environnement ». INRAE a pour ambition d'être un acteur clé des transitions nécessaires pour répondre aux grands enjeux mondiaux.

Face à l'augmentation de la population et au défi de la sécurité alimentaire, au dérèglement climatique, à la raréfaction des ressources et au déclin de la biodiversité, l'institut a un rôle majeur pour construire des solutions et accompagner la nécessaire accélération des transitions agricoles, alimentaires et environnementales.

la science pour la vie, l'humain, la terre



www.inrae/presse